KTH Electrical Engineering

Exploiting Quantized Channel Norm Feedback Through Conditional Statistics in Arbitrarily Correlated MIMO Systems

IEEE TRANSACTIONS ON SIGNAL PROCESSING
Volume 57, Issue 10, Pages 4027-4041, October 2009.

EMIL BJÖRNSON, DAVID HAMMARWALL, AND BJÖRN OTTERSTEN

Stockholm 2009

ACCESS Linnaeus Center
Signal Processing Lab
Royal Institute of Technology (KTH)

DOI: 10.1109/TSP.2009.2024266
KTH Report: IR-EE-SB 2009:010

Exploiting Quantized Channel Norm Feedback Through Conditional Statistics in Arbitrarily Correlated MIMO Systems

Emil Björnson, *Student Member, IEEE*, David Hammarwall, *Member, IEEE*, and Björn Ottersten, *Fellow, IEEE*

***Abstract*—In the design of narrowband multi-antenna systems, a limiting factor is the amount of channel state information (CSI) available at the transmitter. This is especially evident in multi-user systems, where the spatial user separability determines the multiplexing gain, but it is also important for transmission-rate adaptation in single-user systems. To limit the feedback load, the unknown and multi-dimensional channel needs to be represented by a limited number of bits. When combined with long-term channel statistics, the norm of the channel matrix has been shown to provide substantial CSI that permits efficient user selection, linear precoder design, and rate adaptation. Herein, we consider quantized feedback of the squared Frobenius norm in a Rayleigh fading environment with arbitrary spatial correlation. The conditional channel statistics are characterized and their moments are derived for both identical, distinct, and sets of repeated eigenvalues. These results are applied for minimum mean square error (MMSE) estimation of signal and interference powers in single- and multi-user systems, for the purpose of reliable rate adaptation and resource allocation. The problem of efficient feedback quantization is discussed and an entropy-maximizing framework is developed where the post-user-selection distribution can be taken into account in the design of the quantization levels. The analytic results of this paper are directly applicable in many widely used communication techniques, such as space-time block codes, linear precoding, space division multiple access (SDMA), and scheduling.**

***Index Terms*—Channel gain feedback, estimation, MIMO systems, norm-conditional statistics, quantization, Rayleigh fading, space division multiple access (SDMA).**

I. Introduction

WIRELESS communication systems with antenna arrays at both the transmitter and receiver have the ability of greatly improving the capacity over single-antenna systems. The potential gains have been shown for narrowband channels in [1] and [2], under the assumption of independent and identically distributed zero-mean complex Gaussian channel coefficients between the transmit and receive antennas. Such channels are often referred to as uncorrelated Rayleigh fading, since there is no correlation in the spatial dimension and the envelope of the received signal is Rayleigh distributed. From a mathematical point of view, uncorrelated Rayleigh fading channels occur naturally when the antenna separation is large and the scattering in the propagation channel is sufficiently rich. However, it has been shown experimentally that the channel coefficients are often spatially correlated in outdoor scenarios [3], and correlation frequently occurs in indoor environments as well [4], [5]. This motivates the analysis of the more general case of Rayleigh fading where the channel coefficients are arbitrarily correlated.

Channel variations are normally characterized by small-scale and large-scale fading [6]. The former describes changes in the signal paths of the order of the carrier wavelength and is time- and frequency-dependent. To avoid the frequency dependency we consider narrowband block-fading channels; that is, the channel matrix is constant for a block of symbols and then updated independently from the assumed Gaussian distribution for the next block. The large-scale fading corresponds to variations in the channel statistics due to effects like shadowing by buildings and power decay due to propagation distance. These effects are typically frequency independent and slowly varying in time. Hence, the transmitter and receiver can keep track on the statistics by reverse-link estimation or a negligible feedback overhead.

In single-user multiple-input multiple-output (MIMO) systems, the small-scale fading can be mitigated with using orthogonal space-time block codes (OSTBCs) [7]–[9]. Using only statistical channel state information (CSI) at the transmitter, the capacity can be unexpectedly good if linear precoding takes care of the spatial correlation [9]–[12]. In practice, a small amount of channel gain feedback is however necessary for rate adaptation to achieve this performance. In multi-user MIMO systems the situation is somewhat different, because the multi-user diversity gain depends on the amount of instantaneous CSI available at the transmitter [13], [14]. This CSI can be exploited to schedule users for transmission on time-frequency slots and spatial directions in which they experience particularly strong gains. Unfortunately, the amount of feedback needed to achieve full CSI is prohibitive in many realistic scenarios. Therefore, the design of

Manuscript received August 16, 2008; accepted May 03, 2009. First published June 02, 2009; current version published September 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhi Tian. Parts of this work is supported in part by the FP6 project Cooperative and Opportunistic Communications in Wireless Networks (COOPCOM), Project Number: FP6-033533. This work was previously presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, March 30–April 4, 2008 and the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Cannes, France, September 15–18, 2008.

E. Björnson is with the Signal Processing Laboratory, ACCESS Linnaeus Center, Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden (e-mail: emil.bjornson@ee.kth.se).

D. Hammarwall is with Ericsson Research, SE-164 80 Stockholm, Sweden (e-mail: david.hammarwall@ericsson.com).

B. Ottersten is with the Signal Processing Laboratory, ACCESS Linnaeus Center, Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden, and also with the securityandtrust.lu, University of Luxembourg, L-1359 Luxembourg-Kirchberg, Luxembourg (e-mail: bjorn.ottersten@ee.kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2024266

limited feedback systems that capture most of the performance has been an active research topic.

Many multi-user limited feedback systems are based on linear precoding. Although this approach is only asymptotically optimal in the number of users [15], the loss in performance comes with a substantial decrease in complexity compared with non-linear precoding (e.g., optimal dirty-paper coding [16]). One approach to linear precoding in space division multiple access (SDMA) is to allocate users to a set of beams based on feedback of their achieved channel gains. These beams can either be generated randomly [14] or belong to a fixed grid of beams [17]. Another approach is to design and adapt the precoder matrix to statistical user information and feedback of instantaneous CSI. This can be implemented in a zero-forcing fashion [18]–[20], where the co-user interference is made zero (for full CSI) or statistically small and manageable (for partial CSI). Although this strong zero-forcing condition is suboptimal, it provides a simple design structure and can achieve close-to-optimal performance if the amount of feedback is correctly scaled with the signal-to-interference-and-noise ratio (SINR) [18]. In general, the type of approach that is most favorable depends on various system parameters, such that coherence time, number of users, spatial correlation, and average SINR.

Feedback of quantized gain information plays an important role in the design of both user-selection algorithms and linear precoders. In [21], channel norm based user-selection was shown to provide close-to-optimal performance asymptotically in the number of transmit antennas. When considering zero-forcing precoding and limited feedback, it was proposed in [18] that each user should feed back its normalized channel vector using a codebook and calculate a regular zero-forcing precoder. Additional feedback of the instantaneous channel norm is however required to estimate the SINR and perform reliable rate adaptation [22]. In spatially correlated systems, the long-term statistics provide directional information and feedback of the channel norm is sufficient to perform efficient statistical zero-forcing [19] and estimate the instantaneous SINR that is used for rate adaptation [23]. In neither of these papers, channel gain quantization or multi-antenna receivers are considered. With multiple antennas at both sides, more degrees of freedom are available in the interference cancellation, but the precoder and receiver combining design problem becomes considerably more difficult. Some of these problems were addressed in [20].

Herein, we analyze the impact of channel gain information on Rayleigh fading MIMO systems with arbitrary spatial correlation. The conditional statistics and minimum mean square error (MMSE) framework derived in [23] for correlated systems with single-antenna users are generalized to cover general fading environments, multi-antenna users, and quantized gain information. The contributions to communication are an entropy-maximizing quantization framework that can be applied to gain feedback and the derivations of closed-form estimators of the instantaneous SINR in single- and multi-user systems, using such gain feedback. These results can be applied to handle gain feedback and rate adaptation in system both with and without additional feedback of directional channel information.

Notations

For notational convenience we use boldface (lower case) for column vectors, $\mathbf{x}$, and (upper case) for matrices, $\mathbf{X}$. With $\mathbf{X}^T$, $\mathbf{X}^H$, and $\mathbf{X}^*$ we denote the transpose, the conjugate transpose, and the conjugate of $\mathbf{X}$, respectively. The Kronecker product of two matrices $\mathbf{X}$ and $\mathbf{Y}$ is denoted $\mathbf{X} \otimes \mathbf{Y}$, $\mathrm{vec}(\mathbf{X})$ is the column vector obtained by stacking the columns of $\mathbf{X}$, and $\mathrm{diag}(x_1, \ldots, x_N)$ is the N-by-N diagonal matrix with $x_1, \ldots, x_N$ at the main diagonal. If the ijth element of a matrix $\mathbf{X}$ is x_{ij}, then $[\mathbf{X}]_{ij} = x_{ij}$. The distribution of circularly symmetric complex Gaussian vectors is denoted $\mathcal{CN}(\bar{\mathbf{x}}, \mathbf{R})$, with mean value $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{R}$.

The notation $\triangleq$ is used for definitions. The squared 2-norm of a vector $\mathbf{x}$ is denoted $\|\mathbf{x}\|^2$ and the squared Frobenius norm of a matrix $\mathbf{X}$ is denoted $\|\mathbf{X}\|^2$, and both are defined as the sum of the squared absolute values of all the elements. The sum of absolute values of all the elements in $\mathbf{x}$ is denoted $\|\mathbf{x}\|_1$. If $\mathcal{S}$ is a set, then the set members are denoted $\mathcal{S}(1), \ldots, \mathcal{S}(|\mathcal{S}|)$, where $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$.

Let $\mathbf{x} = [x_1, \ldots, x_n]^T$. The generalized Heaviside step function $H_a(\mathbf{x})$ is 1 if $x_i \geq 0$ for all i and $\sum_{i=1}^n x_i \geq a$, and 0 otherwise. The function $H_{a,b}(\mathbf{x})$ is 1 if $x_i \geq 0$, for all i, and $a \leq \sum_{i=1}^n x_i < b$, and 0 otherwise. Finally, $\delta(x)$ denotes Dirac's delta function.

A. *System Model*

Consider the downlink of a communication system with a single base station equipped with an array of n_T antennas and several mobile users, each with an array of n_R antennas. The symbol-sampled complex baseband equivalent of the narrowband flat-fading channel to user k is represented by $\mathbf{H}_k \in \mathbb{C}^{n_R \times n_T}$. The elements of $\mathbf{H}_k$ are modeled as Rayleigh fading with arbitrary correlation, and thus we assume that $\mathrm{vec}(\mathbf{H}_k) \in \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$. The received vector $\mathbf{y}_k(t) \in \mathbb{C}^{n_R}$ of user k at symbol slot t is modeled as

$$\mathbf{y}_k(t) = \mathbf{H}_k \mathbf{x}(t) + \mathbf{n}_k(t) \tag{1}$$

where the vector of transmitted signals is denoted $\mathbf{x}(t) \in \mathbb{C}^{n_T}$ and the power of the system is normalized such that $\mathbf{n}_k(t) \in \mathbb{C}^{n_R}$ is white noise with elements that are distributed as $\mathcal{CN}(0, 1)$.

The system model in (1) depends on three different time scales. The variations in the matrix $\mathbf{H}_k$ are modeled by quasi-static block-fading; that is, the channel realization is constant for a block of symbols and then modeled as independent in the next block. Within a block, only the noise $\mathbf{n}_k(t)$ and the transmitted signal $\mathbf{x}(t)$ are changing. The statistics change very slowly, measured in the number of blocks, and it is therefore assumed that the current correlation matrix $\mathbf{R}_k$ is known to both the base station and user k.

B. Feedback-Based Estimation of Weighted Channel Norms

To achieve reliable rate estimation and exploit the spatial and multi-user diversity, the transmitter often needs more information than just the channel statistics. Such partial and instantaneous CSI can be estimated at the receiver side and then fed back to the transmitter [24]. When the channel conditions change rapidly with time, the number of feedback symbols spent on achieving partial CSI not only reduces the time the information can be used at the transmitter before it is outdated but also the number of symbols available for data transmission on the reverse link. Hence, the feedback needs to represent some limited amount of information that can be described efficiently by a small number of bits.

In a block-fading environment, the feedback system can in principle be described as a cyclical system that estimates and feeds back partial CSI in the beginning of each block to improve the system performance during the rest of the block. The results herein are however not limited to this type of fading. For simplicity, we assume that there exists an error-free feedback channel from each mobile user to the base station.

The instantaneous CSI can be divided into directional information and gain information, herein the latter will be considered. Throughout this paper, we consider the estimation of weighted squared Frobenius norms of the channel at the transmitter [20], [23], where the weights are known at the transmitter but not necessarily at the receiver. On the contrary, the channel is only perfectly known to the receiver and any instantaneous CSI exploited at the transmitter must be conveyed over the limited feedback link. The generic estimation problem that we focus on is

$$\begin{aligned}&\text{Estimate} \quad \|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^2\\ &\quad\text{given} \quad \rho \triangleq \|\widetilde{\mathbf{H}}\|^2 \text{ or a quantized version } a \leq \rho < b. \end{aligned} \quad (2)$$

In this formulation, we have the weighting matrix $\mathbf{B} \in \mathbb{C}^{n_B \times n_C n_D}$ and the effective channel $\widetilde{\mathbf{H}} \triangleq \mathbf{CHD}$, where $\mathbf{C} \in \mathbb{C}^{n_C \times n_R}$ and $\mathbf{D} \in \mathbb{C}^{n_T \times n_D}$ are matrices known to the receiver. In the area of communication, two interesting feedback and estimation scenarios can be formulated in terms of the generic problem.

1) The receive combiner matrix and precoder matrix are known to the receiver and are used as $\mathbf{C}$ and $\mathbf{D}$, respectively. The squared norm of the effective channel $\widetilde{\mathbf{H}} = \mathbf{CHD}$ is fed back to the transmitter. This information is used to estimate the weighted squared norm $\|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^2$, which is either the total channel gain ($\mathbf{B} = \mathbf{I}$) or the gain in a certain spatial subspace.
2) Either the receive combiner matrix, the precoder matrix, or both matrices are unknown to the receiver at the time of feedback. In these cases, the effective channel becomes $\widetilde{\mathbf{H}} = \mathbf{HD}$, $\widetilde{\mathbf{H}} = \mathbf{CH}$, or $\widetilde{\mathbf{H}} = \mathbf{H}$, respectively, and the squared norm $\|\widetilde{\mathbf{H}}\|^2$ is fed back. This information is used to estimate the weighted squared norm $\|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^2$, where $\mathbf{B}$ may represent receive combiner and/or precoder matrices that are known to the transmitter.

The results of this paper are independent of the quantization, but a quantization framework is proposed in Section III and adapted to multi-user systems in Section IV-B.

C. Outline

In Section II, we analyze the special case of feedback of $\rho_k = \|\mathbf{H}_k\|^2$ with a diagonal correlation matrix $\mathbf{R}_k$. Closed-form expressions of the conditional moments of the elements in $\mathbf{H}_k$ are derived for both exact norm feedback and a quantized norm. A short overview of the applications of these results in renewal theory is provided. In Section III, the results are generalized for communication purposes. A general entropy-maximizing quantization framework is presented and the results of Section II are used to characterize the distribution of the effective squared channel norm and to derive an MMSE estimator of weighted squared norms, given quantized norm information. Section IV shows how these results are applicable on MMSE estimation of signal/interference powers and rate adaptation in single- and multi-user systems. Some of the results are illustrated numerically in Section V and conclusions are drawn in Section VI.

II. Analysis of Zero-Mean Complex Gaussian Vectors With Norm Information

In this section, we consider an N-dimensional vector $\mathbf{v} = [v_1, \dots, v_N]^T \in \mathbb{C}^N$, for $N \geq 1$, with zero-mean and independent complex Gaussian entries—that is, $\mathbf{v} \in \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$. First, the distribution of the squared norm $\rho \triangleq \|\mathbf{v}\|^2$ will be presented. Then, expressions of the pth order conditional moment and (p_i, p_j)th order conditional cross-moment are derived for the cases of either an exactly known norm ρ or a known interval $a \leq \rho < b$ (representing a quantization of $\|\mathbf{v}\|^2$). These moments will be used in Section IV to derive a MMSE estimator of weighted squared norms as formulated in (2), and their corresponding mean squared errors (MSEs).

Without loss of generality, we assume that the diagonal elements, $\lambda_i > 0$, of the positive definite correlation matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$ are ordered such that elements with identical distributions have adjacent indices. When analyzing $\|\mathbf{v}\|^2$, we distinguish between three different cases, depending on the distinctness of the variances λ_i (hereafter called eigenvalues):

- identical eigenvalues: $\lambda_1 = \cdots = \lambda_N = \mu_1$, for some μ_1;
- distinct eigenvalues: $\lambda_i \neq \lambda_j$, for all $i \neq j$;
- one or more sets of repeated eigenvalues among λ_j.

While the former two cases are clearly structured and commonly treated in literature, the third case needs further specification [25]. Let the M distinct values among the eigenvalues be $\mu_1, \dots, \mu_M$ $(1 \leq M \leq N)$, with the strictly positive multiplicities $r_1, \dots, r_M$ $(\sum_{i=1}^{M} r_i = N)$. Then, we have the characterization

$$\begin{aligned}\lambda_1 = \cdots = \lambda_{r_1} &= \mu_1\\ &\vdots\\ \lambda_{r_1+\cdots+r_{M-1}+1} = \cdots = \lambda_{r_1+\cdots+r_M} &= \mu_M. \end{aligned} \quad (3)$$

To simplify the notation, we gather the eigenvalue multiplicities in a vector $\mathbf{r} = [r_1, \dots, r_M]^T$ and define the function $r(j) = m$ that gives the group index of λ_j from $\{1, \dots, M\}$

(i.e., $1 \leq m \leq M$ is the integer that satisfies $\sum_{i=1}^{m-1} r_i < j \leq \sum_{i=1}^{m} r_i$).

These three cases are directly applicable to systems with uncorrelated fading (identical eigenvalues), correlated fading (distinct eigenvalues), and Kronecker-structured systems (see Section III) with correlation at either the transmitter or receiver (repeated eigenvalues with either multiplicity n_R or n_T).

Next, the probability density function (pdf) of the squared norm $\|\mathbf{v}\|^2 \triangleq \sum_{i=1}^{N} |v_i|^2$ will be given for the three cases described above. Since $v_i \in \mathcal{CN}(0, \lambda_i)$ for all i, then $|v_i|^2 \in \mathrm{Exp}(1/\lambda_i)$ and the squared norm $\|\mathbf{v}\|^2$ is the sum of independent exponentially distributed variables (each with the rate $1/\lambda_i$). In the case of identical eigenvalues, the pdf is that of a scaled χ^2_{2N}-distribution (i.e., an Erlang distribution):

$$f_{\|\mathbf{v}\|^2}^{\mathrm{ident}}(\rho) = \frac{\rho^{N-1} e^{-\rho/\mu_1}}{(N-1)!\mu_1^N} H_0(\rho) \tag{4}$$

where $H_0(\rho)$ is the Heaviside step function. In the case of distinct eigenvalues, the pdf of $\|\mathbf{v}\|^2$ is well-known in the field of renewal theory [25] and was derived for communications purposes in [23]:

$$f_{\|\mathbf{v}\|^2}^{\mathrm{dist}}(\rho) = \sum_{k=1}^{N} \frac{e^{-\rho/\lambda_k}}{\lambda_k \prod\limits_{\substack{l=1 \\ l \neq k}}^{N} \left(1 - \frac{\lambda_l}{\lambda_k}\right)} H_0(\rho). \tag{5}$$

In the third case, with repeated eigenvalues that satisfy (3), the pdf was derived in [25] and [26]:

$$f_{\|\mathbf{v}\|^2}^{\mathrm{repeat}}(\rho) = H_0(\rho) \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \sum_{k=1}^{M} \sum_{l=1}^{r_k} \frac{\Psi_{k,l,\mathbf{r}}}{(r_k - l)!} (-\rho)^{r_k - l} e^{-\rho/\mu_k} \tag{6}$$

where

$$\Psi_{k,l,\mathbf{r}} = (-1)^{r_k - 1} \sum_{\mathbf{i} \in \Omega_{k,l}} \prod_{j \neq k} \binom{i_j + r_j - 1}{i_j} \left(\frac{1}{\mu_j} - \frac{1}{\mu_k}\right)^{-(r_j + i_j)} \tag{7}$$

with $\mathbf{i} = [i_1, \ldots, i_M]^T$ from the set $\Omega_{k,l}$ of all partitions of $l-1$ (with $i_k = 0$) defined as

$$\Omega_{k,l} = \left\{ [i_1, \ldots, i_M] \in \mathbb{Z}^M; \sum_{j=1}^{M} i_j = l - 1, i_k = 0, i_j \geq 0 \ \forall j \right\}. \tag{8}$$

One remark is that the pdf in (6) actually becomes that in (4) if $M = 1$ and that in (5) if $M = N$. Since the expressions with identical and distinct eigenvalues are simpler and very useful in practice, we will distinguish between all three cases throughout the paper.

A. Conditional Statistics: Known Norm Value or Interval

Next, we will consider the conditional statistics of the elements of $\mathbf{v}$ when its squared norm $\|\mathbf{v}\|^2$ is known exactly or in a quantized way. The absolute value and the phase of a complex Gaussian variable are independent [16]. Thus, $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$ can be identically expressed as $\mathbf{v} = [|v_1| e^{j\phi_1}, \ldots, |v_N| e^{j\phi_N}]^T$, where the phase ϕ_i is uniformly distributed in $[0, 2\pi)$ and $|v_i|^2 \in \mathrm{Exp}(1/\lambda_i)$ for all i. Observe that information regarding $\|\mathbf{v}\|^2 \triangleq \sum_{i=1}^{N} |v_i|^2$ will not provide any knowledge of the phases. The squared magnitudes of the individual elements, $|v_i|^2$, will however depend on $\|\mathbf{v}\|^2$.

In this section, we will derive closed-form expressions of the pth-order conditional moment of $|v_i|^2$ and (p_i, p_j)th order conditional cross-moment of $|v_i|^2$ and $|v_j|^2$. This will be done in two different cases, namely when the squared norm $\rho \triangleq \|\mathbf{v}\|^2$ is either known exactly or when a quantization is known. We denote the quantized squared norm with ϱ and it represents the information $a \leq \|\mathbf{v}\|^2 < b$, for some real-valued interval parameters. This type of quantized information can, for example, be achieved by feedback. The conditional moments derived in the section will be used in Section IV for MMSE estimation and MSE calculation of weighted squared norms in systems with either perfect or quantized squared norm feedback.

The following theorem gives closed-form expressions of the conditional moments in the case of an exactly known squared norm ρ. Although the expressions are quite simple in their structure, two elementary functions $G^{(1)}_{N,M,i,j}(\rho)$ and $G^{(2)}_{N,M,L,i,j,m}(\rho)$ are introduced to achieve a more convenient presentation. These are defined and discussed in Appendix A. Observe that the mean value of an element is given by $p = 1$, the quadratic mean by $p = 2$, and that $p_i = p_j = 1$ gives the cross-correlation.

Theorem 1 (Conditional Moments With Known Norm): Let $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ has strictly positive eigenvalues λ_i and $N \geq 2$. Define $\rho \triangleq \|\mathbf{v}\|^2$. In the case of identical eigenvalues (i.e., $\lambda_i = \mu_1$ for all i), the pth order conditional moment of $|v_i|^2$ and (p_i, p_j)th order conditional cross-moment between $|v_i|^2$ and $|v_j|^2$ ($i \neq j$) are

$$\begin{aligned} &\mathbb{E}\{|v_i|^{2p} | \rho\} \\ &\quad = \frac{(N-1)}{\rho^{N-1}} (-1)^{N-2} G^{(1)}_{p,N-2,1,1}(\rho), \\ &\mathbb{E}\{|v_i|^{2p_i} |v_j|^{2p_j} | \rho\} \\ &\quad = \begin{cases} \frac{(N-1)(N-2)}{\rho^{N-1}} (-1)^{N-3} G^{(2)}_{p_i,p_j,N-3,1,1,1}(\rho), & N > 2, \\ \frac{(-1)^{p_j}}{\rho} G^{(1)}_{p_i,p_j,1,1}(\rho), & N = 2. \end{cases} \end{aligned} \tag{9}$$

In the case of distinct eigenvalues, the corresponding moments are

$$\begin{aligned} &\mathbb{E}\{|v_i|^{2p} | \rho\} \\ &\quad = \frac{\frac{1}{\lambda_i}}{f_{\|\mathbf{v}\|^2}^{\mathrm{dist}}(\rho)} \sum_{\substack{k=1 \\ k \neq i}}^{N} \frac{e^{-\rho/\lambda_k} G^{(1)}_{p,0,i,k}(\rho)}{\lambda_k \prod\limits_{l \notin \{i,k\}} \left(1 - \frac{\lambda_l}{\lambda_k}\right)}, \\ &\mathbb{E}\{|v_i|^{2p_i} |v_j|^{2p_j} | \rho\} \\ &\quad = \frac{\frac{1}{\lambda_i \lambda_j}}{f_{\|\mathbf{v}\|^2}^{\mathrm{dist}}(\rho)} \begin{cases} \sum\limits_{\substack{k=1 \\ k \notin \{i,j\}}}^{N} \frac{e^{-\rho/\lambda_k} G^{(2)}_{p_i,p_j,0,i,j,k}(\rho)}{\lambda_k \prod\limits_{l \notin \{i,j,k\}} \left(1 - \frac{\lambda_l}{\lambda_k}\right)}, & N > 2, \\ (-1)^{p_j} e^{-\rho/\lambda_j} G^{(1)}_{p_i,p_j,i,j}(\rho), & N = 2. \end{cases} \end{aligned} \tag{10}$$

Finally, if the eigenvalues are nondistinct and nonidentical, let $\tilde{\mathbf{r}} = [\tilde{r}_1, \ldots, \tilde{r}_M]^T$ be the eigenvalue multiplicities when the elements involved in the moments have been removed. The pth order conditional moment of $|v_i|^2$ and (p_i,p_j)th order conditional cross-moment between $|v_i|^2$ and $|v_j|^2$ ($i \neq j$) are

$$\begin{aligned}\mathbb{E}\{|v_i|^{2p}|\rho\} &= \frac{1}{f_{\|\mathbf{v}\|^2}(\rho)} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \\ &\times \sum_{k=1}^{M}\sum_{l=1}^{\tilde{r}_k} \frac{\Psi_{k,l,\tilde{\mathbf{r}}}}{(\tilde{r}_k-l)!} e^{-\rho/\mu_k} G^{(1)}_{p,\tilde{r}_k-l,r(i),k}(\rho) \\ \mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\rho\} &= \frac{1}{f_{\|\mathbf{v}\|^2}(\rho)} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \\ &\times \sum_{k=1}^{M}\sum_{l=1}^{\tilde{r}_k} \frac{\Psi_{k,l,\tilde{\mathbf{r}}}}{(\tilde{r}_k-l)!} e^{-\rho/\mu_k} G^{(2)}_{p_i,p_j,\tilde{r}_k-l,r(i),r(j),k}(\rho).\end{aligned} \tag{11}$$

Proof: The proof is given in Appendix C. ∎

Observe that Theorem 1 only handles the case of $N \geq 2$, but the solution in the special case of $N=1$ is trivial: $|v_1|^{2p} = \rho^p$. The theorem generalizes the previous results of [23], where expressions of the first and second order moment and cross-correlation were derived in the special case of distinct eigenvalues.

Next, we proceed with deriving closed-form expressions of the same conditional moments and cross-moments as in the previous theorem but in the case of quantized norm information. Once again, the expressions contain some functions that are defined in Appendix A.

Theorem 2 (Conditional Moments With Known Norm Interval): Let $\mathbf{v} = [v_1,\ldots,v_N]^T \in \mathcal{CN}(\mathbf{0},\mathbf{\Lambda})$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_N)$ has strictly positive eigenvalues λ_i and $N \geq 2$. Let ϱ contain the quantized information $a \leq \|\mathbf{v}\|^2 < b$ (where $0 \leq a \leq b$). In the case of identical eigenvalues (i.e., $\lambda_i = \mu_1$ for all i), the pth order conditional moment of $|v_i|^2$ and (p_i,p_j)th order conditional cross-moment between $|v_i|^2$ and $|v_j|^2$ ($i \neq j$) are

$$\begin{aligned}\mathbb{E}\{|v_i|^{2p}|\varrho\} &= \frac{1}{g_\varrho^{\mathrm{ident}}} \frac{(-1)^{N-2}}{\mu_1^N (N-2)!} \widetilde{G}^{(1)}_{p,N-2,1,1}(a,b), \\ \mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\varrho\} &= \frac{1}{g_\varrho^{\mathrm{ident}}} \begin{cases} \frac{(-1)^{N-3}}{\mu_1^N (N-3)!} \widetilde{G}^{(2)}_{p_i,p_j,N-3,1,1,1}(a,b), & N>2, \\ \frac{(-1)^{p_j}}{\mu_1^2} \widetilde{G}^{(1)}_{p_i,p_j,1,1}(a,b), & N=2 \end{cases}\end{aligned} \tag{12}$$

where

$$g_\varrho^{\mathrm{ident}} = \sum_{j=0}^{N-1} \frac{\left(\frac{a}{\mu}\right)^j e^{-a/\mu} - \left(\frac{b}{\mu}\right)^j e^{-b/\mu}}{j!}. \tag{13}$$

In the case of distinct eigenvalues, the corresponding moments are

$$\mathbb{E}\{|v_i|^{2p}|\varrho\} = \frac{\frac{1}{\lambda_i}}{g_\varrho^{\mathrm{dist}}} \sum_{\substack{k=1\\k\neq i}}^{N} \frac{\widetilde{G}^{(1)}_{p,0,i,k}(a,b)}{\lambda_k \prod\limits_{l \notin \{i,k\}} \left(1-\frac{\lambda_l}{\lambda_k}\right)},$$

$$\mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\varrho\} = \frac{\frac{1}{\lambda_i\lambda_j}}{g_\varrho^{\mathrm{dist}}} \begin{cases} \sum\limits_{\substack{k=1\\k\notin\{i,j\}}}^{N} \frac{\widetilde{G}^{(2)}_{p_i,p_j,0,i,j,k}(a,b)}{\lambda_k \prod\limits_{l\notin\{i,j,k\}} \left(1-\frac{\lambda_l}{\lambda_k}\right)}, & N>2, \\ (-1)^{p_j} \widetilde{G}^{(1)}_{p_i,p_j,i,j}(a,b), & N=2 \end{cases} \tag{14}$$

where

$$g_\varrho^{\mathrm{dist}} = \sum_{k=1}^{N} \frac{e^{-a/\lambda_k} - e^{-b/\lambda_k}}{\prod\limits_{i\neq k}\left(1-\frac{\lambda_i}{\lambda_k}\right)}. \tag{15}$$

Finally, if the eigenvalues are nondistinct and nonidentical, let $\tilde{\mathbf{r}} = [\tilde{r}_1,\ldots,\tilde{r}_M]^T$ be the eigenvalue multiplicities when the elements involved in the moments have been removed. The pth order conditional moment of $|v_i|^2$ and (p_i,p_j)th order conditional cross-moment between $|v_i|^2$ and $|v_j|^2$ ($i \neq j$) are

$$\begin{aligned}\mathbb{E}\{|v_i|^{2p}|\varrho\} &= \frac{1}{g_\varrho^{\mathrm{repeat}}} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \\ &\times \sum_{k=1}^{M}\sum_{l=1}^{\tilde{r}_k} \frac{\Psi_{k,l,\tilde{\mathbf{r}}}}{(\tilde{r}_k-l)!} \widetilde{G}^{(1)}_{p,\tilde{r}_k-l,r(i),k}(a,b) \\ \mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\varrho\} &= \frac{1}{g_\varrho^{\mathrm{repeat}}} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \\ &\times \sum_{k=1}^{M}\sum_{l=1}^{\tilde{r}_k} \frac{\Psi_{k,l,\tilde{\mathbf{r}}}}{(\tilde{r}_k-l)!} \widetilde{G}^{(2)}_{p_i,p_j,\tilde{r}_k-l,r(i),r(j),k}(a,b)\end{aligned} \tag{16}$$

where

$$\begin{aligned}g_\varrho^{\mathrm{repeat}} &= \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \sum_{k=1}^{M}\sum_{l=1}^{r_k} \frac{\Psi_{k,l,\mathbf{r}}}{(-\frac{1}{\mu_k})^{r_k-l+1}} \\ &\times \sum_{j=0}^{r_k-l} \frac{(\frac{b}{\mu_k})^j e^{-b/\mu_k} - (\frac{a}{\mu_k})^j e^{-a/\mu_k}}{j!}.\end{aligned} \tag{17}$$

Proof: The proof is given in Appendix C. ∎

This section will be concluded by Theorem 3 that gives the MMSE estimator of $\rho \triangleq \|\mathbf{v}\|^2$ from the quantized information $a \leq \rho < b$. Observe that the theorem completes Theorem 2 for $N=1$.

Theorem 3 (Norm Estimation From Known Norm Interval): Let $\mathbf{v} = [v_1,\ldots,v_N]^T \in \mathcal{CN}(\mathbf{0},\mathbf{\Lambda})$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_N)$ has strictly positive eigenvalues λ_i. Let $\rho \triangleq \|\mathbf{v}\|^2$ and let ϱ contain the quantized information $a \leq \rho < b$ (where $0 \leq a \leq b$). The conditional pth order moment of $\|\mathbf{v}\|^2$, given ϱ, is

$$\mathbb{E}\{\rho^p|\varrho\} = \frac{1}{g_\varrho^{\mathrm{ident}}} \frac{\widetilde{G}^{(3)}_{N+p-1,1}(a,b)}{\mu_1^N (N-1)!}, \tag{18}$$

$$\mathbb{E}\{\rho^p|\varrho\} = \frac{1}{g_\varrho^{\mathrm{dist}}} \sum_{k=1}^{N} \frac{\widetilde{G}^{(3)}_{p,k}(a,b)}{\lambda_k \prod\limits_{j\neq k}\left(1-\frac{\lambda_j}{\lambda_k}\right)}, \tag{19}$$

and

$$\mathbb{E}\{\rho^p|\varrho\} = \frac{1}{g_\varrho^{\text{repeat}}} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \sum_{k=1}^{M} \sum_{l=1}^{r_k} \frac{\Psi_{k,l,\mathbf{r}} \widetilde{G}_{r_k+p-l,k}^{(3)}(a,b)}{(-1)^{r_k-l}(r_k-l)!} \tag{20}$$

when the eigenvalues are identical (i.e., $\lambda_i = \mu_1$ for all i), distinct, or neither identical nor distinct, respectively. The variables g_ϱ^{ident}, g_ϱ^{dist}, and $g_\varrho^{\text{repeat}}$ are given in (13), (15), and (17), respectively.

Proof: The proof is given in Appendix C. ■

In the remaining sections, the analytic results of Theorem 1, 2, and 3 will be applied to problems in wireless communications. The results of this section are however general and have important applications in other areas, for example in the analysis of N-out-of-M systems with exponential failure rates in renewal theory [25], [27]. In principle, these systems consist of M components and the system will keep running until N of them have break down. The time between the $(i-1)$th and ith component failure is distributed as $\text{Exp}(1/\lambda_i)$ (i.e., failures may change the failure rates of the surviving components). Thus, $\|\mathbf{v}\|^2$ is the time to system failure. The results herein can be used for MMSE estimation of the time between component failures, given the exact time of system failure or a time interval (e.g., if the functionality is tested only at certain occasions). Similarly, the MSE and the correlation between component failures can be calculated, and the time of system failure can be MMSE estimated, given a time interval.

III. Norm Feedback and MMSE Estimation of Weighted Squared Channel Norms

In this section, we return to the generic estimation problem in (2) and the system model in (1). Thus, the effective channel used for norm feedback is $\widetilde{\mathbf{H}}_k = \mathbf{C}_k\mathbf{H}_k\mathbf{D}_k$, where $\text{vec}(\mathbf{H}_k) \in \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$ and $\mathbf{C}_k, \mathbf{D}_k$ are arbitrary matrices known at the receiver. In this section, we will first develop a general entropy-maximizing quantization framework. Then, the results of Section II will be used to derive the distribution of the squared norm $\rho_k = \|\widetilde{\mathbf{H}}_k\|^2$ of the effective channel, which is necessary to apply the quantization framework to the problem at hand. Finally, we solve the estimation problem in (2) by deriving the MMSE estimator, and its MSE, of the weighted squared norm $\|\mathbf{B}_k \text{vec}(\widetilde{\mathbf{H}}_k)\|^2$, conditioned on exact or quantized feedback of ρ_k. As described in Section I-B, the weighting matrix $\mathbf{B}_k$ can represent receive combining and precoding matrices. The applications of this section on user-selection, link-adaptation, and linear precoding will be considered in Section IV. The user index will be dropped in this section for brevity.

The results herein are derived for a general positive semi-definite correlation matrix $\mathbf{R}$, but we will also give the corresponding expressions in the special case of Kronecker-structured correlation. In this widely used model, the transmit and receive side correlation are separable as $\mathbf{R} = \mathbf{R}_{\text{Tx}}^T \otimes \mathbf{R}_{\text{Rx}}$, where $\mathbf{R}_{\text{Tx}} \in \mathbb{C}^{n_T \times n_T}$ and $\mathbf{R}_{\text{Rx}} \in \mathbb{C}^{n_R \times n_R}$ are the positive semi-definite transmit and receive correlation matrices, respectively. As a result, the matrix $\mathbf{H}$ can in this case be decomposed as

$$\mathbf{H} = \mathbf{R}_{\text{Rx}}^{1/2}\mathbf{G}\mathbf{R}_{\text{Tx}}^{1/2} \tag{21}$$

where the elements of $\mathbf{G}$ are independent and identically distributed (i.i.d.) as $\mathcal{CN}(0,1)$. The eigenvalues of $\mathbf{R}$ become the products of any two eigenvalues of $\mathbf{R}_{\text{Tx}}$ and $\mathbf{R}_{\text{Rx}}$, respectively. Depending on the amount of spatial correlation at the transmitter and receiver, the eigenvalues of $\mathbf{R}$ are either identical (e.g., if $\mathbf{R}_{\text{Tx}} = \mathbf{R}_{\text{Rx}} = \mathbf{I}$), distinct (e.g., if distinct eigenvalues at both sides), or consist repeated eigenvalues (e.g., when one of the sides is spatially uncorrelated with either $\mathbf{R}_{\text{Rx}} = \mathbf{I}$ or $\mathbf{R}_{\text{Tx}} = \mathbf{I}$). Eigenvalues that are measured in practice are naturally distinct, but clustering of those that are close-to-equal may be necessary to achieve numerical stability. Recall that these three cases correspond to those in Section II.

A. General Entropy-Maximizing Quantization Framework

Next, we will present a general framework for quantization of a stochastic variable $X \in [0,\infty)$, with the cumulative distribution function (cdf) $F(x)$, for the purpose of finite rate feedback. This variable may represent the weighted squared norm of a communication system, but the results are valid for any continuous cdf that fulfills $F(x)=0$, for $x \leq 0$, and $F(x) < 1$, for $x < \infty$.

With quantization, we mean the process of dividing a continuous range of values into a finite number of intervals. Herein, we consider L-bits quantization of the range $[0,\infty)$ of X, which means that the range is divided into 2^L disjoint intervals $[a_{i-1}, a_i)$, $1 \leq i \leq 2^L$. In our context, the purpose of the quantization is feedback and storage of the variable using L bits. Note that each interval, $\mathcal{Q}$, should be seen as a representative for all values of the original variable that lies in the interval. The actual value in the interval that best represents the quantized information, $\mathcal{Q}$, will change depending on the application (e.g., estimation of X or some function of it). When designing the quantization, we need to choose the decision boundaries a_i, for $0 \leq i \leq 2^L$, so that some design criteria is fulfilled. There is no over-all optimal criteria, but from an information-theoretical perspective it makes sense to maximize the entropy of the quantization and thereby the average amount of channel information that is fed back.

Lemma 1 (Entropy-Maximizing Quantization): Let X be a stochastic variable with a continuous cdf $F(x)$, that fulfills $F(x)=0$, for $x \leq 0$, and $F(x) < 1$, for $x < \infty$. Assume that the sample space, $[0,\infty)$, of X is quantized into 2^L disjoint intervals ($1 \leq i \leq 2^L$), where the ith interval is $[a_{i-1}, a_i)$ with $a_0 = 0$ and $a_{2^L} = \infty$. The interval boundaries that maximizes the entropy of X are given by

$$a_i = F^{-1}\left(\frac{i}{2^L}\right), \quad i = 1, \ldots, 2^L - 1. \tag{22}$$

This quantization will make the outcome of X equally probable in all the quantization intervals.

Let $\mathcal{Q} \in \{1,\ldots,2^L\}$ denote the index such that the outcome $X \in [a_{\mathcal{Q}-1}, a_{\mathcal{Q}})$. The quantization maximizes the mutual